

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2001-216195

(P2001-216195A)

(43) 公開日 平成13年8月10日 (2001.8.10)

(51) Int.Cl.⁷
G 0 6 F 12/08識別記号
5 0 7
5 2 3
5 5 7F I
G 0 6 F 12/08

7-コード* (参考)

5 0 7 E
5 2 3 E
5 5 7

審査請求 未請求 請求項の数15 O L (全 7 頁)

(21) 出願番号 特願2000-397620 (P2000-397620)

(22) 出願日 平成12年12月27日 (2000. 12. 27)

(31) 優先権主張番号 0 9 / 4 7 3 7 8 1

(32) 優先日 平成11年12月29日 (1999. 12. 29)

(33) 優先権主張国 米国 (U S)

(71) 出願人 591236448

エスティーマイクロエレクトロニクス、イ
ンコーポレイテッドS T M i c r o e l e c t r o n i c s ,
I n cアメリカ合衆国、 テキサス 75006,
カーロルトン、 エレクトロニクス ドラ
イブ 1310

(72) 発明者 ランス レズリー フレイク

アメリカ合衆国、 コロラド 80303,
ボルダー、 オマハ プレイス 5473

(74) 代理人 100057793

弁理士 小橋 一男 (外1名)

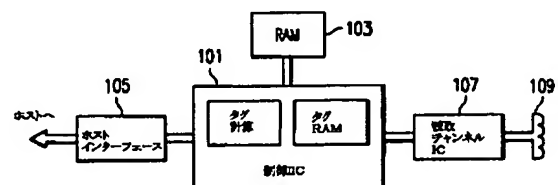
最終頁に続く

(54) 【発明の名称】 分数的二進寸法型キャッシュ

(57) 【要約】

【課題】 キャッシュ使用と非キャッシュ使用との間で二進寸法型メモリ空間を効率的に共用することを可能とするキャッシュを形態特定し且つアクセスする技術を提供する。

【解決手段】 複数のブロックを有する格納装置を設け、各ブロックをブロックアドレスで識別する。アクセス要求がターゲットブロックアドレスを識別する。ターゲットブロックアドレスは上部部分と下部部分とを有している。非二進割算は上部部分に関して実施され、商と余りとを発生する。その余り部分は下部部分と結合されてインデックスを作成する。そのインデックスはタグメモリ構造に適用されて該タグメモリ構造における1つ又は1組のエントリを選択する。選択されたエントリの内容が商部分と比較されて、そのターゲットブロックがキャッシュにおいて表されているか否かを決定する。



あって、更に詳細には、メモリ装置のフラクション即ち分数を使用してキャッシュする方法及びシステムに関するものである。

【0002】

【従来の技術】データ処理システムはデータプログラムコードを格納するために多様なデータ格納（記録）機構に依存している。各格納機構は関連するレイテンシ即ち待ち時間を有しており、それは格納装置へデータを書込む場合及びそれからデータを読取る場合に発生する遅延を意味している。格納機構は、例えば物理的にデータ処理要素近くに位置しているスタックランダムアクセスメモリ（SRAM）等の低い待ち時間の機構のものからSRAMよりも数桁大きな待ち時間を有している磁気、光学的及び遠隔格納機構にわたるものが存在している。大量格納装置は物理的に且つ論理的にデータプロセッサ近くに位置している作業メモリよりも一層大きな待ち時間を有する傾向がある。

【0003】設計のコスト及び複雑さを著しく増加させることなしに性能を向上させることが可能な技術に対する必要性が常に存在している。キャッシュはデータ格納（記憶）システムの性能を改善するために使用される1つの技術である。キャッシュ技術は磁氣的及び光学的ディスク格納装置のような大量格納装置に関連する待ち時間を隠すものである。キャッシュ技術は、選択したプログラム情報、そうでなければより高い待ち時間の格納装置内に格納されるメモリアドレス又はデータを保持する比較的低い待ち時間のメモリを提供する。キャッシュ技術はキャッシュ機構内へコピーされるデータのどの部分を選択するために格納されているデータ内にしばしば存在する空間的及び時間的な参照の局所性の原理を利用している。処理要素によって必要とされるデータのコピーがキャッシュ内に存在する限り、データプロセッサは低い待ち時間の装置と関連する遅延の影響を受けるのみであり、従って性能が著しく改善される。

【0004】多くのデータ処理システム、例えば、埋込型システムでは、システムの全てのメモリ条件に対して単一の物理的メモリ装置を使用する。このことが行われる理由は、市販されているメモリ集積回路（IC）が全ての機能を提供するために充分な使用可能なメモリ容量を有しているからである。マルチチップを使用することは非効率的である。然しながら、単一のメモリ装置の断片的な部分に対してこれらの全く別の機能を割り当てることには問題がある。

【0005】集積回路（IC）メモリは、性質上、二進寸法型インクリメント（たとえば、 $2^{16}=64\text{Kbit}$ 、 $2^{24}=1\text{Gbit}$ ）で格納容量を規定する。他の部分を非キャッシュ動作のために予約しておきながらメモリICの一部をキャッシュとして割り当てることには特定の問題が存在している。1例として、ディスクドライブ即ちディスク駆動装置はファームウェアテーブル及

びコンフィギュレーション（形態）情報を保持するためにメモリを使用するが、これらは従来のメモリIC容量の一部のみを必要とするに過ぎない。残りのメモリ容量はディスクアクセス時間を改善するためにより遅い磁氣的又は光学的格納装置からのデータをキャッシュするために割り当てることが望ましい。従来技術においては、ICメモリ装置の一部のみをキャッシュに対して効率的に割り当てることに困難性が存在していた。

【0006】従来のシステムは1つ又はそれ以上のセグメントをキャッシュに割り当てるために「セグメント化」したメモリアーキテクチャを使用している。各セグメントは循環型のものとして構成することが可能である。適応型セグメント化技術はセグメントの数及び寸法を動的に集積することを可能とする。これらの技術は単一のメモリ装置を実効的にキャッシュ使用と非キャッシュ使用との間で共用することを可能とする。然しながら、セグメント化したアーキテクチャは複雑な制御論理を実現することを必要とする。更に、セグメント化したメモリは、しばしば、従来のタグメモリ制御型キャッシュアーキテクチャよりも性能が劣ったものとなることがある。

【0007】タグメモリ制御型キャッシュ技術は主に汎用コンピュータシステム用に開発されたものであり、その場合に、そのメモリ管理は複数個の集積回路チップを使用して行われる。従来、データアドレスはタグ部分とインデックス部分とに分割される。タグ部分はメモリアドレスの所定数の最大桁ビットを有しており且つインデックス部分は所定数の最小桁ビットを有している。

【0008】キャッシュライン即ちキャッシュブロック（キャッシュの最も小さなアドレス可能な部分）がデータで充填されると、該ターゲットアドレスのインデックス部分は充填すべく使用可能な1個又はそれ以上の組のキャッシュブロックを識別する。識別された組内の1つのキャッシュブロックが選択され且つそのデータは選択されたキャッシュブロック内に書込まれ、一方そのタグ部分は該キャッシュブロックと関連しているタグメモリ内に書込まれる。データがメモリから必要とされると、そのインデックス部分を使用してそのデータを包含する可能性がある1つ又はそれ以上の組のキャッシュブロックを識別する。識別された組に対するタグメモリがサーチされて、一致するタグ値がその中に格納されているかを決定する。一致が存在する場合には、データをキャッシュから読み出すことが可能であり且つメインメモリアccessが回避される。

【0009】タグ部分とインデックス部分とにタグアドレスを分割することは、実効的には、二進割算処理であるが、それを達成するために何等計算上の資源を必要とするものではない。ターゲットアドレスの下部ビットから作成されたインデックスは二進寸法型タグメモリ、従って、二進寸法型キャッシュをカバーする。メモリシス

取られるべきブロックアドレスは、本明細書においては、「ターゲットブロックアドレス」と呼称される。制御IC101は特別目的集積回路であって、典型的に、応用特定集積回路(ASIC)又はプログラマブルゲートアレイ(PGA)装置として供給される。制御IC101はディスク媒体からデータを読み取り且つディスク媒体へデータを書込むためにホストインターフェース105からのコマンドに応答する制御信号を発生すべく論理を実行する。

【0018】これらの動作は、読取/書込ヘッド109をターゲットブロックと整合させ、回転媒体の回転速度を制御し、且つその他の本発明の完全なる理解のためには重要ではない同様の制御動作を包含している。ディスク媒体上の使用不可能なブロックを考慮するために、制御IC101は使用不可能な即ち損傷された位置をトラッキングし且つ要求を使用不可能な位置から使用可能な位置へマッピングする。これらの制御機能はある量のメモリが制御IC101に対して使用可能であることを必要とする。このメモリはコンフィギュレーション(形態特定)情報、パラメータ情報、欠陥ブロックマップ等をロードするために使用することが可能である。これらのタイプのメモリの使用は、メモリ103のキャッシュ特徴の一部ではないメモリの例である。上述した如く、本発明は上述したキャッシュ使用及び非キャッシュ使用の間でのメモリ103の共用のための効率的なメカニズムを可能なものとさせる。

【0019】本発明によれば、制御IC101が、図2により詳細に示したように、タグRAM及びタグ計算ユニットを有している。本発明に基づくメカニズムはタグ管理型キャッシュ構成を可能なものとさせ、その場合にキャッシュエントリはRAM103内に格納される。タグRAM201は複数個のエントリを有しており、各エントリは現在キャッシュされているブロックアドレスを表す「タグ」を保持している。各キャッシュエントリに対して1つのタグエントリが存在している。

【0020】本発明は、各ブロックアドレスを単一キャッシュエントリにおいてのみキャッシュすることが可能であることを意味する直接マップ型キャッシュとして構成されている。直接マップ型構成においては、キャッシュエントリが第一ブロックアドレスによって使用されており且つ第二ブロックアドレスをキャッシュせねばならない場合には、第一キャッシュアドレスは第二キャッシュアドレスに対して場所を与えるためにキャッシュから追い出されねばならない。動作において、この追い出しルーチンを修正することが可能な多様なキャッシュ置換アルゴリズムが存在している。本発明は、予測可能な複雑性における増加と共に連想的キャッシュ構成へ拡張することが可能である。連想的構成はある適用例において性能を改善することが可能である。

【0021】理解すべきことであるが、タグメモリ20

1内に格納されるタグ値は従来のタグ値と異なっている。前述した如く、従来のタグ管理型キャッシュはタグ値としてサブセットのブロックアドレス(例えば、上部ビットの予め選択した数)を使用する。それと対照的に、本発明はタグ計算論理202を使用してタグ値を計算する。本発明によれば、タグ値は予め選択した数の上部ビットから派生されるが、これらのビットとは異なっている。

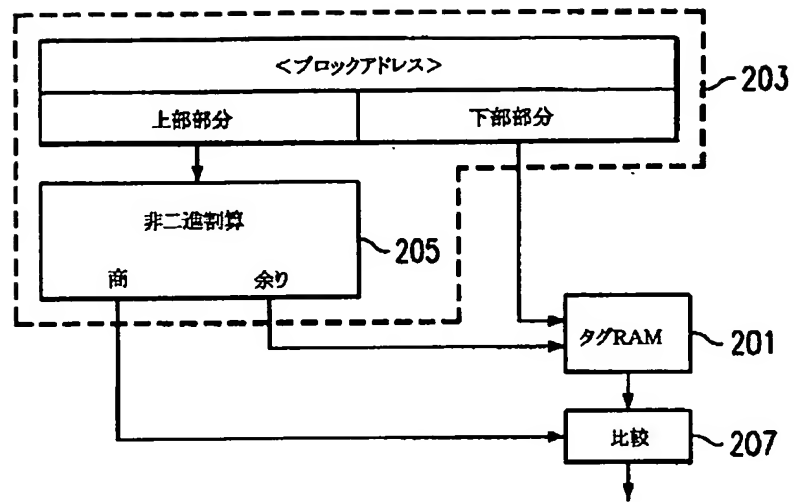
【0022】図2の例においては、例えば28ビットからなるブロックアドレスが使用される。特定の適用例の必要性を満足するためにより多くの又はより少ない数のビットを使用することが可能である。特定の例においては、ブロックアドレスは上部部分と下部部分とを有しているものとみなされる。上部部分及び下部部分の寸法は全体的なブロックアドレスに適用される割算器205の非二進割算演算期間中に値を変化させることのないビットが下部部分内に包含されるように画定される。

【0023】一方、上部部分と下部部分との間の区別は無視することが可能であり且つ全体的なブロックアドレスを割算ユニット105を介して処理することが可能である。割算演算と関連する待ち時間及び複雑性は、通常、割算されるビットの数の関数であるので、本発明のこの特徴は、下部部分ビットが割算演算をバイパスし且つ改善した性能を提供することを可能としている。

【0024】特定の例として、メモリ103が2MB二進寸法型メモリ装置を有しているシステムについて検討する。非キャッシュ目的のために割り当てられている区域は128KB(即ち、1/16)であり、一方キャッシュ目的のために割り当てられている区域は1.875MB(即ち、15/16)である。本発明によれば、タグ計算論理が機能してブロックアドレスを15の倍数によって割算する。各々が512バイトのデータブロックの場合、1.875MBキャッシュは3840個(15の倍数)のエントリを保持することが可能である。従って、直接マップ型キャッシュの場合には、タグインデックスはブロックアドレスを3840で割算した後の12ビットの余りとして得られ、且つタグメモリ内に格納される値が商である。簡単で最適なものではない具体例は28ビットのブロックアドレス全部を非二進割算ユニット205へ供給し、それが3840による割算を実施して17ビットの商と12ビットの余りとを発生する。

【0025】図2に示したより最適な実施例においては、非二進割算ユニット205が、15での割算演算を実施した後に17ビットの商と4ビットの余りとを発生する項として20ビット上部部分を受付ける。タグインデックスの残りの8ビットは直接的に下部部分ビットから取られる。この実施例は3840による割算処理によって変化されるものではないビットであるという認識の下に、8個の下部部分ビットを割算ユニット205を介して処理することを回避している。

【図2】



フロントページの続き

(72)発明者 ティモシー リチャード フェルドマン
 アメリカ合衆国、コロラド 80027,
 ルイスビル、グラント アベニュー
 1029